DRAFT SUMMARY OF HINDRANCES TO INCREASED ENERGY EFFICIENCY IN SERVERS AND DATA CENTERS

Compiled in Support of the Conference on Enterprise Servers and Data Centers: Opportunities for Energy Savings
January 31-February 1, 2006 – Sun Microsystems and AMD Conference Facilities

PURPOSE OF THIS DOCUMENT: This document is intended to: 1) recognize the considerable work that has been aimed at identifying both hindrances to efficiencies in servers and the facilities that house them and opportunities for overcoming these hindrances; and 2) serve as a starting point for a working session focused on improving efficiencies within data centers, to be held on the first day of the title conference.

SOURCES OF INFORMATION: The majority of the concepts relayed in this summary were gleaned from the following reports;

- "Design Recommendations for High-Performance Data Centers." Rocky Mountain Institute. February 4-5, 2003. Provides a system-wide set of recommendations for server design to optimize energy usage.
- Shamshoian, Gary *et al*. "High-Tech Means High-Efficiency: The Business Case for Energy Management in High-Tech Industries." US Dept. of Energy. Contains a detailed quantitative analysis of the cost savings possible though energy-efficient technology.
- "Server Power Supplies." *High Performance Buildings: Data Centers*. Lawrence Berkeley National Laboratories. December 2005. Provides detailed testing of a group of server power supplies. Includes general comments and conclusions derived from the test data and further actions that may be necessary to gain a better understanding of power supply optimization.
- Tschudi, William, *et al*. "Roadmap for Public Interest Research for High-Performance Data Centers." Introduces the RMI report listed above and provides a detailed process to implement greater energy efficiency in data centers.

BACKGROUND: Servers continue to become more pervasive and more powerful, a combination that has greatly increased energy consumption by these devices. Energy inefficiency places a financial burden both on customers – paying to maintain ideal operating conditions – and on manufacturers – paying to develop methods to push the thermal limits of component materials as the increased heat generated by server components climbs higher. "Design Recommendations for High-Performance Data Centers" reports that each additional watt of server power has a value of at least four to eight dollars per watt. This document also states that for every watt being consumed by a computer, roughly two to three additional watts are being drawn from the utility to cool the computer and provide it with protected power.

Despite the desirability of energy savings, a number of unique hindrances can provide difficulty in achieving this goal. Listed below are some examples of these challenges and the solutions to which could help data centers limit costs and maintain reliability. Following the listing of hindrances is an initial list of opportunities for overcoming these hindrances that have been identified by various stakeholders to date.

HINDRANCES

ENERGY BENCHMARKING

- **Benchmark Development**:
  - Performance designations on hardware elements infrequently stress "performance per watt," or similar designations highlighting energy use.
  - Many benchmarks relating performance to energy use are not widely applied.
- **Average Load Considerations**:
  - It is difficult to determine average power loads on servers as a group due to the unique nature of individual workloads and the specific processing required.

EQUIPMENT

- **Uptime vs. Idle-Mode Energy Savings**:
  - The ideal server is a machine with 100% uptime and an instant response time to an information request. This ideal does not mesh well with many current computer energy saving modes, often based on an idle processor or system standby.
  - Current standby methods are implemented only at the personal computer or workstation level, not with the server as a whole.
  - Network traffic can prevent systems from entering an idle mode.
- **"Bloatware**:"
  - With the rapid progression of computing power, programmers do not have as much financial incentive to optimize code as they do to release it to market. If the hardware will allow it to run just as well – transparent to the user – there is no reason to make a program run more efficiently. This results in a computational and an energy waste.
- **Architecture-Specific Requirements**:
  - Specific processor architectures require varying amounts of power to function; this is a secondary consideration to computational functionality for most customers.
- **Unused System Capacity**:
  - Processing capacity and processing load are often not in agreement, resulting in inefficiency while the server is underutilized and fully powered.
- **No Guidelines for Energy Savings:**
  - Guidelines for energy savings in enterprise servers are greatly needed to foster industry-wide acceptance.
- **"Nameplate" Data**:
  - Technical characteristics provided by manufacturers frequently overstate HVAC load and peak electrical requirements, making it difficult for end users to optimize cooling and supply to minimize energy use.

Power Supplies

- **Inefficiency**:
  - Though typically most efficient at maximum loads, power supplies often run at half capacity due to reliability redundancy. For small office tasks, supply redundancy can limit loads on a single supply even lower – less than 15% of capacity. In the power supply alone, this limited usage results in the waste of as much as half of the power entering a computer.
- **Design**:

- By failing to distinguish between component and system cost, and between first and lifecycle cost, current power estimates used in server design are highly inaccurate.
- Since the average load of a power supply is only around 20–25 percent of the maximum rating, power supplies rarely operate at the full-load condition for which they are usually designed.
- Servers can accept a limited number of power supply form factors, hindering temperature and power optimization through alternative configurations.

FACILITIES AND OPERATIONS

- **Implementation Organization**:
  - Implementing energy efficiency measures in servers requires a significant amount of coordination and the combined effort of design, manufacturing, and business elements.

Cooling

- **Outdated Conceptions About Data Center Cooling:**
  - **Alternatives to air**:
    - -- The vast majority of server cooling is achieved by air convection, presenting a commercial barrier to alternative methods. In contrast to the past, air cooling is not as effective at the temperatures created by modern processor densities; increasing amounts of energy are expended to simply continue air cooling.
    - -- Example: though water can conduct 3,467 times as much heat as the same volume of air and requires an order of magnitude less energy to move a given volume, the outdated preconception of water cooling methods which directly contact the circuitry (adding a failure risk) has limited its implementation.
  - **Running temperature**:
    - -- Guidelines relating to acceptable temperature environments for server operation may be too low and result in unnecessary air conditioning.
- **Physical Design Elements**:
  - Higher failure rates at the top of server racks point to flawed use of space and placement of cooling elements.

OTHER

- **Split Incentives**:
  - Costs related to energy use in servers are not divided among the various uses of the technology. End users are not aware of the direct impact of energy on their cost of operation.
  - Users lower on the supply chain are not given breakdowns of the cost to show the potential for savings in energy cost. Users higher on the chain are often not aware of the impact of changes to power implementation on overall manufacturing cost.

- **Second Tier Priority for Building Owners:**
    - Energy issues can be a secondary concern, at best, for building owners/operators. These owners/operators may also lack information that makes the business case for investing in efficiency.

Opportunities

For the End User/Data Facility Manager

- Inclusion of more efficient components, reducing heat waste and HVAC costs.
- Identification of metrics for assessing the efficiency of servers and data centers.
- Creation/expansion of performance-energy benchmarks. These publicized measurements would clearly demonstrate the costs and benefits of choosing power-efficient servers.
- Completion and sharing of rigorous assessments of steps taken to increase efficiency in a range of scenarios with associated costs, LOE, pros and cons.

For Manufacturers and Other Industry Stakeholders

- Implementation of more accurate nameplate labeling. Designers would have a clearer picture of their design constraints, resulting in faster design, simpler implementation, and more control of the end product.
- Modular power supplies. These would allow designers to scale power supply hardware up or down, based on the desired load on the server.
- Take advantage of untapped market prospects for "green servers." End user desire for energy-efficient servers would create a selling point that is currently ignored.
- Reduction of CPU power demands. This has the peripheral benefit of sidestepping the heat challenge that semiconductor manufacturers face in designing product.
- Reduce heat waste from a system perspective. Designers will not have to devote as much design time or physical server space to HVAC requirements.

Benefits to the Environment

Included below is a list of environmental benefits that could be realized if action is taken today.

- Reduction of load on power generation network. Less expansion of the load on power plants due to server operations means reduced greenhouse gas emissions to the air.
- Modular components. The ability to replace obsolete or failed power supply components piece by piece means less physical waste or lead (PB) in landfills. Users will discard a smaller portion rather than the entire device.
- Decrease heat wasted by servers. Thermal contamination of the surrounding environment reduced by the decrease of vented heat into the environment.